基于卷积神经网络和贝叶斯分类器的句子分类模型 *

李文宽 [1,2]，刘培玉 [1,2†]，朱振方 [3]，刘文锋 [1,2,4]

(1. 山东师范大学 信息科学与工程学院，济南 250014；2. 山东省分布式计算机软件新技术重点实验室，济南 250014; 3. 山东交通学院 信息科学与电气工程学院，济南 250014; 4. 菏泽学院 计算机学院，山东 菏泽 274015)

摘 要：传统句子分类模型存在特征提取过程复杂且分类准确率较低等不足，利用当下流行的基于深度学习模型的卷积神经网络在特征提取上的优势，结合传统句子分类方法提出一种基于卷积神经网络和贝叶斯分类器的句子分类模型。该模型首先利用卷积神经网络提取文本特征，其次利用主成分分析法对文本特征进行降维，最后利用贝叶斯分类器进行句子分类。实验结果表明在康奈尔大学公开的影评数据集和斯坦福大学情感分类数据集上，所提出的方法优于只使用深度学习的模型或传统句子分类模型。

关键词：深度学习；句子分类；卷积神经网络；主成分分析法；贝叶斯分类器

中图分类号：TP391.1 doi: 10.19734/j.issn.1001-3695.2018.07.0525

Sentence classification model based on convolution neural network and Bayesian classifier

Li Wenkuan[1,2], Liu Peiyu[1,2†], Zhu Zhenfang[3], Liu Wenfeng[1,2,4]

(1. *School of Information Science & Engineering*, *Shandong Normal University*, *Jinan* 250014, *China*; 2. *Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology*, *Jinan* 250014, *China*; 3. *School of Information Science & Electric Engineering*, *Shandong Jiaotong University*, *Jinan* 250014, *China*; 4. *School of Computer Science*, *Heze University*, *Heze Shandong* 274015, *China*)

Abstract: The traditional sentence classification model has many disadvantages such as complex feature extraction process and low classification accuracy. This paper used the advantages of the popular deep learning model based convolutional neural network in feature extraction, combined with the traditional sentence classification method, proposed a sentence classification model based on convolutional neural network and Bayesian classifier. The model first used convolutional neural network to extract text features, and secondly used principal component analysis method to reduce the dimensionality of text features. Finally, Bayesian classifier were used to classify sentences. The experimental results show that on Cornell University's public film review dataset and Stanford Sentiment Treebank dataset, the method proposed in this paper is superior to the model using only deep learning or the traditional sentence classification model.

Key words: deep learning; sentence classification; convolutional neural network; principal component analysis; Bayesian classifier

0 引言

句子分类[1]是自然语言处理（natural language processing, NLP）中的核心任务之一，近年来受到 NLP 领域学者的广泛关注，成为 NLP 的研究热点。句子分类是人们的观点、情绪评价和对实体如产品、服务、问题、事件、主题和属性等态度的计算研究[2]。

目前常用的句子分类方法分为传统的分类算法和当下最为流行的基于深度学习的分类算法。传统的句子分类算法[3]包含朴素贝叶斯分类器（naive Bayes classifier）、支持向量机（support vector machine）与最大熵模型（maximun entropy model）等。朴素贝叶斯分类器[4]通过训练较少数量的训练集文本便可获取模型参数，在许多复杂的现实场景中，其训练速度较快且能保证准确率，但其在文本预处理阶段，由于需要“考虑表情符号+标点符号二次情感提取”等语法模式，使得特征提取过程比较复杂；支持向量机算法[5]将句子分类任务转换为二次型寻优问题，并可得到全局最优解，有效地解决了在神经网络中无法避免的局部极值问题，但其在文本特征表示时，主要通过词频选择文本特征，完全忽略了句子的上下文结构信息；最大熵模型[6]选择文本特征灵活且不需要额外的独立假设或内在约束，但其对语料库的依赖性较强且因计算量大导致训练时间较长。由此可以看出，传统的句子分类方法在文本模型表示和特征选择阶段，存在特征提取复杂、被提取的特征易忽略上下文结构信息以及模型训练时间较长等不足。

随着深度学习在自然语言处理领域的推广[7,8]，深度学习模型在语言建模和句子分类等方面取得了很大的进展。2013 年 Mikolov 提出了 Word2Vec 模型[9]，使语言中的字词转换为计算机可以理解的稠密向量，适合处理局部序列数据，有效地改善了传统文本表示时向量维度极高、过于稀疏且词与词之间无关联等不足；2014 年 Kim 提出将卷积神经网络应用到句子分类任务中[10]，在数据集较大的情况下，卷积神经网

收稿日期：2018-07-16；**修回日期：**2018-08-27 **基金项目：**国家自然科学基金资助项目（61373148）；国家自然科学基金青年资助项目（61502151）；山东省社科规划项目（17CHLJ18，17CHLJ33，17CHLJ30）；山东省自然科学基金资助项目（ZR2014FL010）；山东省教育厅基金资助项目（J15LN34）

作者简介：李文宽（1993-），男，山东德州人，硕士研究生，主要研究方向为自然语言处理；刘培玉（1960-），男（通信作者），山东潍坊人，教授，博导.主要研究方向为自然语言处理与网络信息安全（liupy@sdnu.edu.cn）；朱振方（1980-），男，山东临沂人，副教授，硕导，主要研究方向为自然语言处理；刘文锋（1978-），男，博士研究生，主要研究方向为自然语言处理.

络不用人工提取特征，简化了传统句子分类算法中复杂的文本特征提取步骤。2017 年 Ma 等人提出基于多层注意力机制的卷积神经网络，并将该网络应用在句子建模上[11]，使其能够更好地捕获局部文本特征，验证了注意力机制和卷积神经网络结合的有效性。由此可以看出，深度学习模型在文本特征的提取和表达方式方面优于传统句子分类方法所采用的方式。

本文将当下流行的基于深度学习模型的卷积神经网络与传统句子分类方法朴素贝叶斯分类器相结合，利用了卷积神经网络提取文本特征辨识度高的优点，同时结合朴素贝叶斯分类器在复杂现实情景中训练速度快、准确率高且参数易获取的优势，准确地提取到文本特征，并有效地解决了句子分类问题。

1　相关工作

1.1　卷积神经网络

1.1.1 *卷积层和池化层*

卷积神经网络(convolutional neural network,CNN)是图像处理领域研究热点之一，它可以直接将多维的原始图像输入网络，避免对图像的复杂预处理过程并提取高辨识度的图像特征[12]。在 NLP 领域中，同样可以在文本特征提取的过程中利用卷积神经网络在图像预处理上的优势，简化对文本预处理的过程，提取到高辨识度的文本特征，减少特征工程的工作量。与传统神经网络不同的是卷积神经网络在输入层和全连接层之间添加了卷积层和池化层，有效地解决了传统神经网络模型参数较多和网络层数限制等问题。卷积层[13] (Convolution Layer)为特征提取层，通过利用不同卷积核对上一层输入矩阵进行卷积操作，提取局部特征组成卷积核特征矩阵，卷积操作计算表达式如式(1)所示：

$$c_j^l = f\left(_{i\in M_j} c_i^{l-1} *? W_j^l + b_j^l\right),? \qquad (1)$$

其中:f 表示非线性激活函数，c_j^l 表示第 l 层的第 j 个特征图，M_j 表示特征图的某一个元素，W_j^l 表示第 l 层的第 j 个特征图卷积核矩阵，b_j^l 表示对应的偏置项。池化层[14](pooling layer)对卷积层的特征向量图进行下采样操作，利用特征图的局部相关原理，在相邻小区域内进行聚合统计，进一步提取更重要的特征信息。同时，不同长度的句子输入可通过池化层生成固定维度的特征向量，并将池化层输出传给全连接层进行分类。目前常用的下采样操作有均值池化(Average Pooling)、最大值池化(Max Pooling)和随机池化(Stochastic Pooling)等。池化操作的计算表达式如式(2)所示。

$$c^{pooling} = flatten(c_{3,1}^{pooling}, c_{3,2}^{pooling}, \ldots,$$
$$c_{3,m}^{pooling}, c_{5,1}^{pooling}, c_{5,2}^{pooling}, \ldots, c_{h,m}^{pooling})? \qquad (2)$$

其中: $c^{pooling}$ 表示池化后输出的特征集合，h 表示不同大小的卷积核，m 表示每组卷积核的数目。

1.1.2 Attention *机制*

注意力机制(attention mechanism)最早应用在图像处理领域，使神经网络能够在处理图像时关注某些重点信息[15]。Attention 机制可理解为从大量信息中有选择地筛选出重要信息并聚焦到这些重要信息上，同时忽略不重要信息，聚焦的过程体现在权重系数的计算上，权重越大越聚焦其对应的信息值。在 NLP 任务中，注意力机制主要作用在抽取文本的语义上，假设某一句子 S 中单个词向量表示为 p_i，在某具体任务中对该句子中单个词的关注度为α_i，则该句子 S 的表示如式(3)所示：

$$V_s = _{i=1}^{|S|}{}_i *? p_i,? \qquad (3)$$

其中:$|S|$ 表示句子按序输入的单词数，通过 $_i$ 的取值不同来改变对句子 S 的关注点。设任务由语义向量 Vq 表示，则 $_i$ 表示为 p_i 和 V_q *的函数* $F(p_i, V_q)$，常用的函数 $F(p_i, V_q)$ 计算方式有向量的张量积[16]和双线性函数[17]。

1.2　主成分分析法

主成分分析法(Principal Component Analysis,PCA)是一种基于统计思想的降维方法[18]。PCA 通过线性变换将原始数据变换为一组各维度线性无关的表示，用于提取数据的主要特征分量，以达到高维数据降维的目的。PCA 基本原理如下:

设原始数据集表示为 $X_{m\times n}$ 矩阵，将 $X_{m\times n}$ 的每一行进行零均值化，计算表达式如式(4)所示，其中 x_{ij} 表示矩阵 $X_{m\times n}$ 第 i 行第 j 列的元素，$\overline{x_i}$ 表示矩阵 $X_{m\times n}$ 第 i 行的均值，S_i 表示矩阵 $X_{m\times n}$ 第 i 行的标准差；

$$x_{ij} = \frac{x_{ij} - \overline{x_i}}{S_i},? \qquad (4)$$

根据式(5)求出协方差矩阵 C，其中 m 表示样本个数，同时求出协方差矩阵 C 的特征值$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k$ 及对应特征向量 d_1，d_2，$\cdots$，d_k；

$$C = \frac{1}{m-1} _{k=0}^{j} (x_j - \bar{x})(x_j - \bar{x})^T,? \qquad (5)$$

根据式(6)求出特征贡献率，并将特征向量按照对应特征值大小从上到下按行排列成矩阵 P，$Y=PX$ 即为降维后的 k 维矩阵。

$$e_i = \frac{_i}{_1^k{}_i},? \qquad (6)$$

1.3　朴素贝叶斯分类器

朴素贝叶斯分类器是基于贝叶斯定理与特征条件独立假设的分类方法[19]。它主要是根据先验概率分布预测样本属于某一类别的后验概率，选取概率最大的类别作为预测类别。整个朴素贝叶斯分类器的分类过程为：对于给定的训练数据集，首先基于特征条件独立假设，学习输入/输出的联合概率分布；然后基于此概率分布，对于给定的输入 x，利用贝叶斯定理求出后验概率最大的输出 y。朴素贝叶斯分类器的数学模型可以表示如下：

假设输入特征向量 $X(x_1, x_2, \cdots, x_n)$为待分类样本，输出空间为类标记集合 $Y=\{c_1, c_2, \cdots, c_m\}$。对样本 X 进行分类，则需要计算 $P(c_1|X), P(c_2|X), \ldots, P(c_m|X)$，那么 X 的预测类别表达式如式(7)所示。

$$P(c_k|X) = ?max\{P(c_1|X), P(c_2|X),$$
$$\ldots, P(c_m|X)\},? \qquad (7)$$

c_k 即为朴素贝叶斯分类器预测的待分类样本类别。

计算式(7)中条件概率的步骤如下：

a)构造已知类标记的训练样本集合；

b)统计训练集中各个特征在每个类别中的条件概率，如 $P(x_1|C_1), P(x_2|C_1), \ldots, ?P(x_n|C_1)$；

c)假定各个特征属性相互独立，则根据贝叶斯定理可得条件概率表达式为式(8)所示。

$$P(c_i|X) = ?\frac{P(X|c_i)P(c_i)}{P(X)},? \qquad (8)$$

d)在式(8)中，分母 $P(X)$对所有类别都是相同的，因此只需将分子最大化即可，化简后的条件概率表达式为

$$P(X|c_i)P(c_i)=?P(c_i)\prod_{i=1}^{m}P(x_i|c_i),? \tag{9}$$

2　模型构建

本文针对词向量的情感词构造词性注意力特征矩阵，在卷积神经网络的基础结构上，提出基于 attention 机制的双通道卷积神经网络，并将此网络与 PCA 和朴素贝叶斯分类器结合，实现句子分类。本节重点介绍模型的整体框架、基于 attention 机制的卷积神经网络具体实现细节和朴素贝叶斯分类器对文本特征的分类过程。

2.1　模型框架

基于卷积神经网络和贝叶斯分类器的句子分类模型的主要思想是通过卷积神经网络提取文本特征，利用 PCA 对提取到的文本特征进行降维，最后利用朴素贝叶斯分类器进行句子分类。模型框架如图 1 所示。

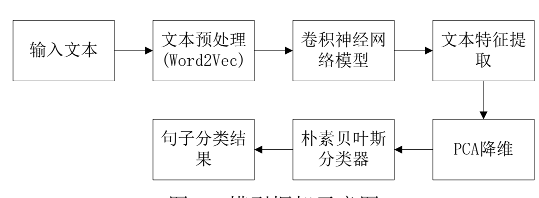


图 1　模型框架示意图

Fig. 1　Model frame schematic

由图 1 可知，算法主要分为两个阶段：训练阶段和测试阶段。在训练阶段，利用 word2vec 表示的训练集文本集合对卷积神经网络模型进行训练，然后将训练好的网络模型应用到所有文本特征的提取，将提取到的高维文本特征进行 PCA 降维处理之后，用降维后的文本特征集合训练朴素贝叶斯分类器。在测试阶段，对测试集文本集合进行预处理后，通过卷积神经网络提取文本特征，通过 PCA 降维，最后将降维的文本特征传入朴素贝叶斯分类器即可得到句子分类结果。

2.2　基于 attention 机制的双通道卷积神经网络

本文提出的基于卷积神经网络和贝叶斯分类器的句子分类模型中，卷积神经网络模型为基于 attention 机制的双通道卷积神经网络，如图 2 所示，此模型共有 5 层。在文本预处理阶段，利用 word2vec 工具将数据集中的句子转换为词向量矩阵，并将词向量矩阵与情感词向量结合形成词性注意力矩阵，两个矩阵构成卷积神经网络的双通道；在卷积神经网络第 2 层（卷积层），卷积窗口在句子长度上进行滑动，一个窗口生成一个值，生成一个长度为 sentence_length+window_size-1 的向量，以此获取输入文本的局部特征；在卷积神经网络的第 3 层（池化层），以滑动窗口的形式进行 average pooling 对卷积层的文本特征矩阵进行池化并保持特征相对位置不变；在卷积神经网络的第 4 层（卷积层），对于上一层池化层特征矩阵，同样以卷积窗口的形式在句子长度上滑动，使文本特征得到更深层次的提取；在卷积神经网络的第 5 层（池化层），对于上一层卷积层特征矩阵，在句长维度上进行 average pooling，得到代表输入句子特征的列向量作为卷积神经网络的输出层；最后通过 PCA 降维处理卷积神经网络的输出层，利用降维后的句子特征训练分类器。

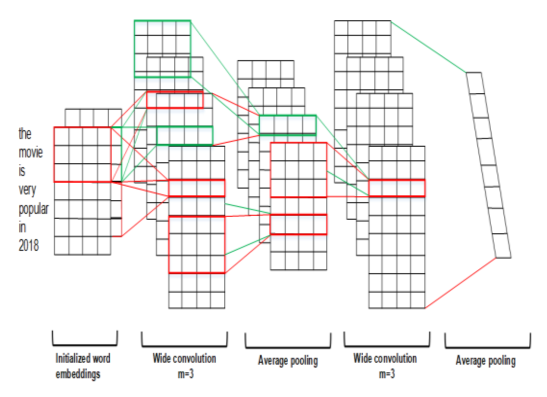


图 2　卷积神经网络模型

Fig. 2　Convolutional neural network model

Attention 机制是在神经网络的各个部位对输入或输出进行加权，这种加权会利用词向量本身或外部的其他向量来强调特征矩阵中相对重要的信息。本文采用词性注意力机制结合词向量形成词性注意力特征矩阵，与词向量输入矩阵结合形成双通道，作为卷积神经网络的输入层。仅依赖文本结构信息进行句子分类会造成准确率偏低，针对此问题，本文通过分词并结合情感词典中的情感词，抽取原文句子中的情感词组成词性注意力特征矩阵。对于长度为 n 的句子 $s=\{w_1,w_2,\ldots,a_i,\ldots,w_n\}$，本文将抽取出句子 s 中情感词 a_i 的词向量，即 $a_i\in R^l$，其中 l 表示情感词向量维度且与 Word2Vec 维度相同。将情感词 a_i 的词向量与句子 s 的词向量矩阵做内积运算，可得到对角矩阵 A^c，计算过程如式(10)(11)所示。

$$A=tensorproduct(a_i,s),? \tag{10}$$

$$A^c_{i,i}=\frac{\exp(A_{i,i})}{\sum_{j=1}^{n}\exp(A_{j,j})},? \tag{11}$$

再利用对角矩阵 A^c 与句子 s 的词向量矩阵做点积运算便可得到词性注意力矩阵 z_i^c，计算过程如式(12)所示。

$$z_i^c=dotproduct(s_i,A^c_{i,i}),? \tag{12}$$

卷积运算是卷积神经网络获取输入特征的关键一步，并以特征矩阵的形式保存在网络结构中，特征矩阵的每一个单元与前一层的局部特征相关联，卷积核在局部特征上做卷积操作，最后通过激活函数得到特征矩阵的值。本文卷积操作采用广义卷积运算，也称宽卷积，可使卷积核扫描整个输入端的句子向量，不限制输入层句子长度 s 与卷积核 m 的大小关系且卷积运算的输出不会为空向量，计算表达式如式(13)所示，卷积核输出为 $c'\in R^{s+m-1}$，c 为输入句子的词向量矩阵，卷积核 $k\in R^m$。由此可以看出，广义卷积操作将卷积核的覆盖范围扩大，解除了输入层句子长度 s 与卷积核 m 的大小关系限制，充分考虑了各种可能存在的特征，在最大程度上保证文本特征提取的多样性，同时保证在句子长度不一致时，有效捕捉完整的句子信息。

$$c'_j=k^T c_{j:\ j+m},\ -m+1\le j\le s,? \tag{13}$$

池化层置于卷积层的下一层，对卷积层局部数据进行抽样或聚合，以此来降低卷积层特征矩阵的维度，同时使其不易出现过拟合。本文池化操作采用 Average Pooling，池化层特征矩阵的计算表达式如式(14)所示。

$$x_i^l=f\left(\beta_i^l down\left(x_i^{l-1}\right)+b_i^l\right) \tag{14}$$

其中：$down()$表示池化函数，β 表示乘性偏置，b 表示加性偏置，x_i^l 表示第 l 层 Pooling 的第 i 个特征值。

2.3　文本特征分类器

文本数据在通过卷积神经网络后，经过 PCA 降维处理，最终得到句子的主要特征属性。本文利用朴素贝叶斯分类器对句子特征进行分类。在给定的句子特征向量中，首先利用贝叶斯条件概率公式，如式(15)所示，$P(x_j|c)$为已知句子类别 c 的特征属性 x_j 的概率；再利用贝叶斯公式计算出已知句子属性属于不同句子类别的后验概率，如式(16)所示；最后根据最大后验概率将该句子归结为具有最大后验概率的句子类别，如式(17)所示。

$$P(x_1,x_2,\ldots,x_n|c)=?_{j=1}^{n}P(x_j|c),? \tag{15}$$

$$P(c|x_1,x_2,\ldots,x_n)=?\frac{P(c)_{j=1}^{n}P(x_j|c)}{P(x_1,x_2??\ldots,x_n)},? \tag{16}$$

$$P(c|x_1,x_2,\ldots,x_n)=argmax?P(c)_{j=1}^{n}P(x_j|c), \tag{17}$$

3　对比实验

3.1　实验数据

为验证本文所提出模型的有效性，实验采用康奈尔大学基于影评数据创建的数据集 MRD (movie review data, https://www.cs.cornell.edu/people/pabo/movie-review-data/) 和斯坦福大学情感分类语料库数据集 SST(Stanford sentiment treebank, https://nlp.stanford.edu/sentiment/)，分别用于二分类(negative,positive) 的句子和五分类 (very negative,negative,neutral,positive,very positive)的句子。其中 MRD 由电影评论数据组成，持肯定态度评论占 1000 篇，持否定态度评论占 1000 篇，标注了褒贬极性的句子各 5331 句，标注了主客观标签的句子各 5000 句，在本文实验中，随机抽取 1400 篇作为训练集，400 篇作为测试集，200 篇作为验证集；SST 数据集是 MRD 数据集的扩展，共有 11855 个句子，人工标注了句子类别，8544 句为训练集，2210 句为测试集，1101 句为验证集。本文实验所用数据的统计如表 1 所示，其中训练集、测试集、验证集之比为 7：2：1。

表 1　实验使用数据统计

Table 1　Statistic of the datasets

数据集	训练集	测试集	验证集
MDR	1400	400	200
SST	8544	2210	1101

3.2　超参数设置

词向量的维度 d=300，选择已经预训练的 300 维谷歌 Word2Vec 词向量文件来映射，将未在 Word2Vec 中出现的文中单词通过随机函数映射为[-1,1]的 300 维随机向量；实验中使用双通道对输入矩阵进行卷积操作，卷积核函数为修正线性单元(rectified linear units,ReLu)，训练通过 SGD(stochastic gradient descent)进行，利用 Zeiler 提出的 Adadelta 优化器[20]；其他实验参数如表 2 所示，首先通过经验设置初始参数，其次在实验数据的验证集上设置模型迭代次数为 100，通过观察交叉熵损失函数的变化进行调参，最后选择在实验数据的验证集上性能最好的一组参数作为训练模型的参数输出。

表 2　实验参数设置

Table 2　Hyper parameters of experiment

parameter	parameter description	value
m	window size	3
p	dropout rate	0.5
s	L2 constrain	3
b	mini-batch size	50

通过卷积神经网络提取的文本特征为 240 维，特征维度过高，本文利用 PCA 对卷积神经网络提取的文本特征进行降维处理。为了研究主成分特征值累计贡献率对分类性能的影响，本文分析了不同累计贡献率下基于卷积神经网络和贝叶斯分类器的句子分类模型的分类准确率，分类准确率与特征值累计贡献率的关系图如图 3 所示。

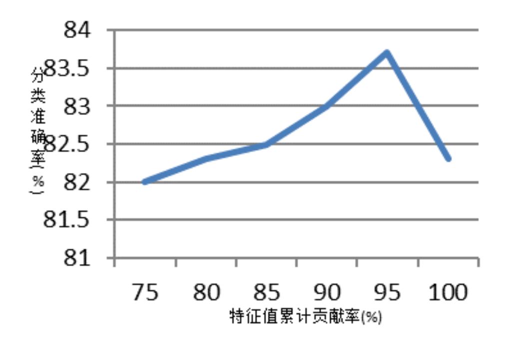


图 3　分类准确率与特征值累计贡献率关系图

Fig. 3　Classification accuracy rate and eigenvalue cumulative contribution rate relationship diagram

由图 3 可知，设特征值累计贡献率为α，当α从 100%减少到 95%过程中，经过卷积神经网络提取到的文本特征经 PCA 降维处理，文本特征中的冗余信息被逐渐剔除，分类准确率逐渐增加；当α=95%时，文本特征中的冗余信息剔除较充分，分类准确率最高；当α从 95%逐渐减小时，部分有用文本特征被剔除，导致分类准确率也随之下降。由此可得，主成分的维数对分类准确率的影响至关重要。在本文实验中，选取α=95%，能够将卷积神经网络提取到的 240 维特征通过 PCA 降维到 80 维，降维效果显著。

3.3　实验结果

为验证本文所提出模型的有效性，实验设置了 4 个 baseline 方法分别为朴素贝叶斯分类器(Naive Bayes Classifier,NBC)、支持向量机(Support Vector Machine,SVM)、卷积神经网络(Convolutional Neural Network,CNN)和 CNN+SVM，在 MRD 和 SST 数据集的测试集上进行对比实验，本文模型与各模型在测试集的实验结果如表 3 所示。

Table 3　Model classification correct rate comparison result

表 3　模型分类正确率对比结果

模型	MDR 正确率(%)	SST 正确率(%)
SVM	77.4	43.9
NBC	76.3	42.8
CNN	81.1	47.4
CNN+SVM	82.3	48.3
CNN+NBC	83.7	49.8

通过实验结果表明，在 MDR 数据集上，基于深度学习模型的 CNN 和 CNN+SVM 的准确率分别为 81.1%和 82.3%，比传统句子分类模型 SVM 提高 3.7%和 4.9%，比传统句子分类模型 NBC 提高 4.8%和 6.0%；在 SST 数据集上，基于深度学习模型的 CNN 和 CNN+SVM 的准确率分别为 47.4%和 48.3%，比传统句子分类模型 SVM 提高 3.5%和 4.4%，比传统句子分类模型 NBC 提高 4.6%和 5.5%。传统句子分类模型 NBC 和 SVM 在数据预处理阶段，重点关注“表情符号+标

点符号二次情感人工标注”等语法模式以及句子的词频特征，但是忽视了上下文文本结构信息，不但复杂化了文本情感特征表示过程且文本情感分类准确率偏低；基于深度学习模型的 CNN 和 CNN+SVM 利用词向量简化了文本情感特征表示过程，并且利用卷积神经网络捕获包含上下文文本结构信息的高质量文本特征，使得深度学习模型句子分类准确率明显高于传统句子分类模型。本文提出的模型在 MDR 数据集上的准确率为 83.7%，比基于深度学习模型的 CNN 和 CNN+SVM 提高 2.6%和 1.4%；在 SST 数据集上的准确率为 49.8%，比基于深度学习模型的 CNN 和 CNN+SVM 提高 2.6%和 1.4%。本文提出的卷积神经网络结构相对了传统卷积神经网络结构，在输入层加入词性注意力机制，使卷积神经网络在文本情感特征学习过程中更关注句子的情感极性目标词；将卷积神经网络提取到的文本情感特征经 PCA 降维处理，剔除掉文本情感特征中的冗余信息；利用朴素贝叶斯分类器训练速度快且在多维特征空间中保持准确率的稳健性，对降维后的文本情感特征进行分类，使得句子分类准确率比基于深度学习模型的 CNN 和 CNN+SVM 的句子分类准确率进一步提升。

图 4 更加形象直观地显示出不同模型在不同数据集上的句子分类准确率。不难看出，当句子分类的类别标签数增加时，传统句子分类模型和深度学习模型的准确率也随之降低，这表明在多分类文本特征提取上，现有的句子分类模型仍不能较准确地提取高识别度的文本特征。

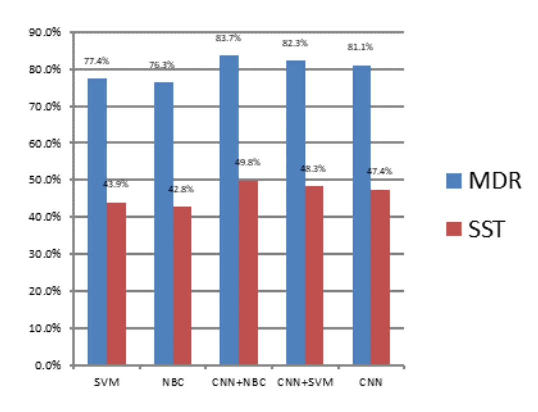


图 4　五种方法在不同实验数据集上地实验结果比较

Fig. 4　Comparison of experimental results of five methods on different experimental data sets

4　结束语

在文本情感倾向性分析任务中，过去的研究主要是改进传统句子分类算法中文本情感特征表示方式和词频统计方式，这类方法存在文本情感特征表示方法复杂且被提取的文本情感特征忽略掉上下文结构信息的缺陷，增加了特征工程的工作量且句子情感分类准确率偏低。本文将基于深度学习模型的卷积神经网络和传统句子分类算法相结合，提出结合改进的卷积神经网络和贝叶斯分类器的句子分类模型，该模型通过引入词向量表示文本情感特征大大降低了文本情感特征表示的工作量。同时，本文提出在卷积神经网络输入层加入词性注意力机制，有助于提升被提取的文本情感特征质量，提高句子情感分类准确率。此外，本文将传统卷积神经网络的输出层转换为贝叶斯分类器，利用贝叶斯分类器训练速度快且在多维特征空间中保持准确率的稳健性以及卷积神经网络在提取上下文结构信息的针对性，通过对比实验验证了将改进的卷积神经网络与贝叶斯分类器相结合有效提高了句子情感分类准确率，并改善了基于深度学习模型的卷积神经网络在输出层黑盒测试的解释性缺陷，增强了深度学习模型在分类过程中的可解释性。

从实验结果可以看出，本文提出的基于卷积神经网络和贝叶斯分类器的句子分类模型对文本情感倾向性多分类效果不理想，所以本文接下来将针对这个问题对卷积神经网络结构进行改进。

参考文献：

[1] Yu Jianfei, Jiang Jing. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification [C]//Proc of Internatiomal Conference on Empirical Methods in Natural Language Processing. 2016: 236-246.

[2] Zhang Lei, Wang Shuai, Liu Bing. Deep learning for sentiment analysis: a survey [J]. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2018,8(4): 1253-1256

[3] Wang Sida, Manning C D. Baselines and bigrams: simple, good sentiment and topic classification [C]//Proc of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2. Stroudsburg, PA: Association for Computational Linguistics, 2012: 90-94.

[4] Jiang Qiaowei, Wang Wen, Han Xu, *et al.* Deep feature weighting in naive Bayes for Chinese text classification [C]//Proc of IEEE International Conference on Cloud Computing and Intelligence Systems. 2016: 160-164.

[5] Joseph L, Zhu Yun, Zhang Yanqing. Support vector machines and Word2vec for text classification with semantic features [C]//Proc of IEEE International Conference on Cognitive Informatics & Cognitive Computing.2015: 136-140.

[6] Hosseini M, Kerachian R. A Bayesian maximum entropy-based methodology for optimal spatiotemporal design of groundwater monitoring networks [J]. Environmental Monitoring & Assessment, 2017, 189 (9): 433.

[7] Kumar A, Irsoy O, Ondruska P, *et al.* Ask me anything: dynamic memory networks for natural language processing [C]/ Proc of the 33rd International Conference on International Conference on Machine Learning. 2016: 1378-1387.

[8] Gatt A, Krahmer E. Survey of the state of the art in natural language generation: core tasks, applications and evaluation [J]. Vestnik Oftalmologii, 2018, 45 (45): 1-16.

[9] Mikolov T, Chen Kai, Corrado G, *et al.* Efficient estimation of word representations in VectorSpace [EB/OL]. [2014-02-10]. http://arxiv. org/pdf/1301. 3781. pdf.

[10] Kim Y. Convolutional neural networks for sentence classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1746-1751.

[11] Ma Dehong, Li Sujian, Zhang Xiaodong, *et al.* Interactive attention networks for aspect-level sentiment classification [C]//Proc of the 26th International Joint Conference on Artificial Intelligence. 2017: 4068-4074.

[12] LeCun Y, Bengio Y. Hinton G. Deep learning [J]. Nature, 2015, 521 (7553): 436.

[13] Jégou S, Drozdzal M, Vazquez D, *et al.* The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation [C]//Proc of IEEE Computer Vision and Pattern Recognition Workshops.2017: 1175-1183.

[14] Fernando B, Gavves E, Oramas M J, *et al.* Rank pooling for action recognition[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39 (4): 773-787.

[15] Firat O, Cho Kyunghyun, Bengio Y. Multi-way, multilingual neural machine translation with a shared attention mechanism [C]//Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 866-875.

[16] Kadlec R, Schmid M, Bajgar O, *et al.* Text Understanding with the Attention Sum Reader Network [C]//Proc of Meeting of the Association for Computational Linguistics. 2016: 908-918.

[17] 北京大学数学系前代数小组. 高等代数[M]. 4 版.北京: 高等教育出版社, 2013 (Pre-Algebra Group, Department of Mathematics, Peking University. Higher algebra [M].4th ed. Beijing: Higher Education Press, 2013)

[18] 李建林. 一种基于 PCA 的组合特征提取文本分类方法 [J]. 计算机应用研究, 2013, 30 (8): 2398-2401. (Li Jianlin. A PCA-based combined feature extraction text classification method [J]. Application Research of Computers, 2013, 30 08): 2398-2401.)

[19] Duan Liguo, Di Peng, Li Aiping. A new naive Bayes text classification algorithm [C]//Proc of IEEE International Conference on Data Mining. 2014: 51-58.

[20] Zeiler M D. ADADELTA: an adaptive learning ratemethod [EB/OL]. [2012-12-22]. http://arxiv. org/pdf/1212. 5701. pdf.